

本周基本确定了论文的整体思路。

上周之前已经对数据管理部分有了基本想法，但是一直纠结于如何在可视查询上无法突破。查阅了近三年的 SIGCHI 的论文，在群体移动数据查询上很少有相关文献。于是退而求其次，猛然发现之前的道路查询其实也可以用于基站轨迹数据。在实际应用中，人群的移动对基站的负荷压力是有影响的。假设人群大多沿着道路移动，那么道路两侧的基站数据能反映出哪些基站负荷大。这对于移动公司设定基站位置和设备功率有一定帮助。另外，研究人群在道路上的移动轨迹对于研究人群的出行方式有一定帮助，例如哪些道路上人流比较密集。结合出租车轨迹数据，完全可以定义一种新的查询模型。以此模型为基础，可以构造从交互到数据组织，再到查询，最后的可视化及分析的完整系统。

系统中的主要问题包括：

- 1、基于道路的可视化查询。这点与先前论文中的查询类似，只是在这里不仅可以查出租车轨迹，而且也可以查道路两旁的基站数据。
- 2、查询模型的定义。抽象为点对象集合在二维空间上的连续移动，需要查询对象集合的移动轨迹。
- 3、查询算法。由于上述查询问题是集合求交，当数据量较大时，难以满足交互的实时性要求，为此需要做两方面的性能提升：
 - 为数据构建局部的时空索引
 - 求交运算在 `spark` 上运算（对集合进行分解，分配到不同的机器上做求交运算，再对结果合并）
- 4、应用案例。分别给出出租车和基站数据的分析结果。

我查阅了 VLDB、SIGMOD 相关论文，移动数据的研究主要集中在数据模型和查询优化上。Gutting 等人主要研究移动数据模型，试图将该模型加入到数据库中。而港科大和新加坡的一群研究人员则主要做查询优化。例如基于道路网约束的轨迹查询、对移动对象建立时空索引、连续空间求交查询等。而对于基于道路的查询模型在数据库领域还没有相关文献。

所以，如果能完成这个系统，那么将会在以下几个方面有所创新：

- 1、基于道路的查询模型
- 2、查询算法及其实现